

ASYNCHRONOUS MESSAGING USING A DYNAMIC ROUTING NETWORK

INVENTORS

TIMOTHY TUTTLE
ROBERT S. DREYER
JEREMY SCOTT LUECK
CHARLES EDWARD CANO
PAUL ANDREW SORENSEN

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit under 35 U.S.C. § 119(e) of U.S. Provisional Application No. 60/256,613, filed December 18, 2000, U.S. Provisional Application No. 60/276,847, filed March 16, 2001, U.S. Provisional Application No. 60/278,303, filed March 21, 2001, U.S. Provisional Application No. 60/279,608, filed March 28, 2001, and U.S. Provisional Application No. 60/280,627, filed March 29, 2001, all of which are hereby incorporated by reference herein.

BACKGROUND

FIELD OF THE INVENTION

[0002] This invention pertains in general to transferring information through digital networks and in particular to transferring information for remotely updating content at client devices through the digital networks.

BACKGROUND ART

[0003] The Internet is a digital network of computers. An individual computer on the Internet is typically identified by an internet protocol (IP) address. A computer on the

Internet sends a packet of information to another computer by routing the packet to a logical port at the destination computer's IP address. The destination computer interprets the packet according to one of several possible protocols determined by the port to which the packet was sent.

[0004] The World Wide Web (the "Web") is a collection of technology and content available on the Internet that allows the content to be routed from server computers to particular destination computers. The Web includes a large number of web pages residing on many different servers. Web pages contain one or more files, or references to one or more files, specifying instructions for presenting the web page and content, such as text, images, applets, video, and/or audio.

[0005] Web pages use a variety of definitional and programming languages to control how information is presented. The most fundamental of these is the Hypertext Markup Language (HTML). HTML uses a system of "tags" to specify how content should be displayed. Recent advances in HTML introduce "style sheets" which help separate content information from display information. HTML has also been modified and extended to provide new capabilities. For example, Extensible Markup Language (XML) adds semantic content to web pages. In addition, Dynamic HTML (DHTML) adds some dynamic content to web pages.

[0006] A web page may also include one or more programs for controlling how the web page is displayed. For example, JAVA[®] applets and JAVASCRIPT[®] scripts may be used to control the display of a web page. In addition, DHTML uses scripts to control the dynamic content. Thus, a web page designer can use applets and scripts to produce

animation effects or modify the display based on user interaction. For example, the designer can write a script that changes the color of a piece of text when a user clicks on a button.

[0007] Devices that display/execute web pages are often called “client devices” or simply “clients.” Client devices include personal computers, web-enabled set-top boxes and televisions, cellular telephones, personal digital assistants and other handheld devices, and special-purpose web-browsing appliances. Client devices typically employ a program called a “web browser” for interpreting the HTML or other display instructions in the web page and displaying the content accordingly. Most web browsers include special functionality, such as a Java Virtual Machine, for executing JAVA[®] applets and/or other applets or scripts embedded in the web pages.

[0008] A client device specifies a web page or other document on the web using a Uniform Resource Locator (URL). A URL has the form “service://server/path/file.” Here “service” refers to the protocol to be used, such as the file transfer protocol (FTP) or the hypertext transport protocol (HTTP). “Server” is the IP address of the server containing the page, and “path/file” specifies the particular web page on the server.

[0009] The Web suffers from a substantial limitation with respect to dynamically updating content in a web page at a client device. The Web’s only mode of operation is for a client device to first request a page from a server and then for the server to send the requested page to the client device. Once the server delivers the page to the client, it typically terminates its connection to the client, and does not retain any information about the client or the page that was sent. For this reason, servers are typically “stateless.” As a

result, client devices drive and control the flow of information around the Web. While client-side control is appropriate in some situations, it does not permit efficient updating of data at the client devices. For example, if a web page contains information that may change, such as the score of a baseball game or a stock quote, the server has no way to inform the client devices that are viewing the page of the change. Instead, the client devices must ask the server for the updated information. However, the client devices do not know when the information on the web page has changed, and thus do not know to ask for the update.

[0010] There are some simple web programming techniques that attempt to update content on client device-side web pages. One approach that web designers use is to rely on the client devices to periodically re-request web pages. This updating can be performed as the result of user action (such as pressing the “refresh” button) or can be automated to occur on a particular schedule (such as by using the HTML Meta Refresh tag to cause the client device to request the page every ‘X’ seconds). Although this technique provides client devices with more up-to-date information, it is very wasteful of resources. In particular, the web server must resend the page even if nothing has changed, and, even when something has changed, it must resend the entire web page rather than just the updated information, which may be only a very small part of the page. Further, attempting to reduce unnecessary requests by decreasing the request rate results in decreasing the currency of the data. This is an unalterable trade off in a client-driven approach.

[0011] The performance of automatic refreshing can be improved somewhat by putting information that may change in a separate frame from information that is less likely to change, and only refreshing the separate frame. A few web designers even write custom JAVA applets to limit refreshing to individual components on a page, such as the score of a soccer game. A willingness to go to such effort illustrates the serious drain of resources caused by frequent refreshing.

[0012] Nevertheless, even custom JAVA applets are not a meaningful attack on this problem. Custom applets require a large separate development effort for each item on each page that might need to be updated. More importantly, most custom applets still update content based upon client-driven requests, although it is possible to design an applet that accepts “pushed” messages. This solution is not scalable to provide updated information for large numbers of client devices and for large numbers of web pages.

[0013] Therefore, there is a need in the art for an efficient way to provide dynamic content to a web page at a client device.

DISCLOSURE OF THE INVENTION

[0014] The above need is met by a dynamic content routing network that routes messages containing data for updating properties of live objects to clients displaying web pages or other representations of data containing the live objects. The web server that initially provides the web pages to the clients does not need to track which clients are currently displaying the live objects. Instead, the information provider or a dynamic content provider (generically referred to as an “input source”) that provided the live

object simply sends an update message to the routing network. The routing network maintains the mappings between the live objects and the clients that are currently displaying them. This routing utilizes bandwidth efficiently because the update messages are provided to the clients only when the live objects change.

[0015] In one embodiment, a web server associated with an input source provides web pages to clients. The web pages contain live objects identified by object IDs. An activation module at the client examines the web page and identifies the object IDs of the live objects on the web page. Then, the activation module establishes a connection with the routing network and registers for the object IDs. The routing network maintains a registry indicating which clients have registered for which live objects.

[0016] An input source provides update messages to the routing network. An update message preferably contains an object ID and data for updating a property of the identified live object. The routing network determines the clients that have registered for the identified object ID, and then routes the message to those clients. The activation modules at the clients receive the message and update the properties of the live object as specified by the data in the message.

[0017] The features and advantages described in this summary and the following detailed description are not all-inclusive, and particularly, many additional features and advantages will be apparent to one of ordinary skill in the art in view of the drawings, specification, and claims hereof.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] FIG. 1 is a high-level block diagram illustrating an environment containing a dynamic content routing network;

[0019] FIG. 2 is an interaction diagram illustrating interactions among a server, information provider, dynamic content provider, client, and routing network to update a property of a live object on a web page;

[0020] FIG. 3 is a high-level diagram graphically indicating the many-to-many mapping performed by the routing network;

[0021] FIG. 4 illustrates two different web pages containing sports scores;

[0022] FIG. 5 is a block diagram illustrating an input source and the tools available to it for generating the update messages;

[0023] FIG. 6 is a flow chart illustrating the steps performed by an embodiment of an activation module;

[0024] FIG. 7 is a block diagram illustrating a lower-level view of the routing network according to an embodiment of the present invention; and

[0025] FIG. 8 is a flow chart illustrating steps performed by a node in a cluster to perform object-based routing of a message received from an input source via the gateway.

[0026] The figures depict an embodiment of the present invention for purposes of illustration only. One skilled in the art will readily recognize from the following

description that alternative embodiments of the structures and methods illustrated herein may be employed without departing from the principles of the invention described herein.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0027] FIG. 1 is a high-level block diagram illustrating an environment 100 containing a dynamic content routing network 110 (hereafter referred to as the “routing network”). The environment 100 also contains a server 112 in communication with a client 114, an information provider 108, and a dynamic content provider 116. Although a typical environment 100 will have hundreds of servers 112 and information providers 108, thousands (or even millions) of clients 114, and multiple dynamic content providers 116, FIG. 1 illustrates only one of each of these entities in order to enhance the clarity of this description.

[0028] The server 112, client 114, information provider 108, dynamic content provider 116, and routing network 110 are preferably in communication via conventional communications links 117 such as those comprising the Internet. The communications links 117 include known wired communications media, such as dedicated or shared data, cable television or telephone lines, and/or known wireless communications media, such as communications over the cellular telephone network using protocols such as the global system for mobile communications (GSM), code division multiple access (CDMA), time division multiple access (TDMA), etc.

[0029] In one embodiment, the entities may each be in communication with one or more Internet Service Providers (ISPs) (not shown) that provide each entity with access to

other computers on the Internet. In addition, the server 112, client 114, information provider 108, dynamic content provider 116, and routing network 110 are preferably each identified by at least one Internet Protocol (IP) address such as "66.35.209.224." The IP address may also have one or more domain names associated with it, such as "bangnetworks.com." Alternative embodiments of the present invention may use alternative addressing schemes and/or naming conventions instead of, or in addition to, those described herein. For example, embodiments wherein one or more of the clients are cellular telephones or other portable devices may rely on different addressing schemes.

[0030] Preferably, the information provider 108 provides web pages or other representations of data to the server 112. The web pages contain one or more "live objects," which are designated to be real-time dynamically-updateable objects. Each live object is identified by an object identifier, or object ID. Preferably, the server 112 provides the web pages 118 to multiple clients 114. The clients 114 contact the routing network 110 and register for update messages for the object IDs on the web page. The routing network 110, in turn, preferably maintains a registry indicating which clients have registered for which object IDs.

[0031] The information provider 108 and/or dynamic content provider 116 send update messages to the routing network 110. These messages can be sent any time the information provider 108 or dynamic content provider 116 wants to update a property of a live object. Each update message preferably identifies a live object and contains data for updating a property of the identified live object. The routing network 110 accesses the registry and determines which clients have registered for the identified object. Then, the

routing network 110 routes the update message to the appropriate clients. Upon receipt of an update message, the clients 114 update the specified property of the live object.

[0032] The routing network 110 provides an efficient one-to-many mapping of objects to clients (and by inference of information, a many-to-many mapping of information providers 108/dynamic content providers 116 to clients) through object-based routing. Messages provided by the information provider 108 and/or dynamic content provider 116 to the routing network 110 are not routed to the clients 114 based entirely on a specified destination; more specifically, they are not routed based on the IP address of the client, as in conventional IP routing schemes. Instead, the messages are routed based on the live objects referenced by the message.

[0033] The mapping and object-based routing provided by the routing network 110 allow the information provider 108 and dynamic content provider 116 to update properties of live objects at a dynamically changing cross-section of clients in real-time, without requiring the information provider or dynamic content provider to track the clients or web pages being viewed by the clients. The clients 114, in turn, do not need to have any a priori knowledge of object IDs – they “discover” which IDs they should register when they receives the web pages 118 from the server 112.

[0034] Object-based routing also allows information providers 108 to dynamically update content on web pages without requiring the clients 114 to re-request the content, and without requiring the information providers 108 or servers 112 to maintain connections with the clients. In this manner, significantly more clients can receive updated content from a given information provider 108 than would be possible utilizing

conventional client-side request-driven transmission control protocol/Internet Protocol (TCP/IP) connections between the clients and the server 112.

[0035] Turning now to the individual entities illustrated in Fig. 1, the server 112 is preferably a conventional computer system configured to act as a web server and serves web pages 118 and other data representations to clients 114. The web pages 118 provided by the server 112 are associated with one or more information providers 108.

[0036] An information provider 108 is an entity providing one or more web pages 118, information contained in web pages, and/or other representations of data served by the server 112. The information provider 108 preferably has a conventional computer system coupled to the Internet. In one embodiment, the server 112 is directly controlled by the information provider 108 (e.g., the server is physically located at the information provider and/or is dedicated to serving only the information provider's web pages). In this embodiment, the server 112 and information provider 108 can be treated as the same entity. In an alternative embodiment, the server 112 serves web pages from multiple information providers.

[0037] As is known in the art, the web pages 118 and other content on the server 112 are specified by uniform resource locators (URLs) having the form "service://server/path/web page." Typically, web pages 118 are obtained via the hypertext transport protocol (HTTP) and thus an exemplary URL for retrieving the web page "b1.html" from the web server having the domain name "www.bangnetworks.com" is "http://www.bangnetworks.com/news/b1.html."

[0038] As used herein, a “web page” is a block of data available from the server 112. In the simplest case, a web page is a file written in the hypertext markup language (HTML). The web page may also contain or refer to one or more other blocks of data, such as other files, text, images, applets, video, and/or audio. In addition, the web page may contain instructions for presenting the web page and its content, such as HTML tags and style sheets. The instructions may also be in the Extensible Markup Language (XML), which is related to HTML and adds semantic content to web pages or the Dynamic HTML (DHTML), which adds some dynamic content to web pages. Additionally, the instructions may take the form of one or more programs such as JAVA[®] applets and JAVASCRIPT[®] and/or DHTML scripts.

[0039] As used herein, the phrase “web page” also refers to other representations of data served by the server 112 regardless of whether these data representations include characteristics of conventional web pages. These data representations include, for example, application programs and data intended for the web browser 120 or other application programs residing at the clients 114 or elsewhere, such as spreadsheet or textual (e.g., word processing) data, etc.

[0040] In a preferred embodiment, objects at the client, such as web pages and elements of web pages, can be designated as “live” by the information provider 108. Properties of a live object can be dynamically updated in real-time at the client 114 by the information provider 108 or another entity acting on behalf of the information provider. As used herein, an “object” is any datum or data at the client 114 that can be individually identified or accessed. Examples of objects include elements of web pages such as text

characters and strings, images, frames, tables, audio, video, applets, scripts, HTML, XML, and other code forming the web page, variables and other information used by applets, scripts and/or code, URLs embedded in the web page, etc. Application and operating system constructs are also objects. For example, cells of spreadsheets, text in word processor documents, and title bars and messages displayed by the operating system or applications are objects. Preferably, multiple objects can be grouped together into a single, logical object. Thus, an object can be defined at any desired or useful level of granularity.

[0041] Since content on a web page is conceptualized and organized by “object,” the present invention essentially abstracts web pages and web page content, and other modules and/or functionality at the client 114, away from the HTML code or other conventional representation. This abstraction allows the information provider 108 to update a property of an object without concern for the location, display format, or other specifics of how the data is being represented at the client 114.

[0042] Live objects have associated “properties” which include any modifiable data related to the object or referenced with respect to the object. The information provider 108 typically, but not necessarily, provides initial settings for the properties of live objects provided to the client 114. The properties may or may not affect the visual representation of the object in the web page or other data representation. A property may affect an internal aspect of the object and, thus, a change to the property may not have any direct effect on a web page containing the object. For example, the property may affect whether particular aspects of the object are modifiable, how the object responds to user

input or other stimuli, etc. Additionally, a property may also have a direct effect on how the object is displayed at the client 114. For example, the property may affect the content, color, typeface, size, formatting, or other attribute of text, images, or other data displayed by the object. Other properties may occupy parts of the spectrum between having no effect on the visible representation of the object and having a direct effect on the visible representation of the object. For example, a web page showing scores of football games may include a list of games and the current scores of the games as of the time the server 112 serves the web page. The list of games, subset of games to be displayed, and the scores of the games can be designated as live objects (or properties of a single live object) and updated as necessary or desired.

[0043] A property can also preferably include instantiating an instance of the object or invoking functionality of the object. For example, a property of a browser window object may include functionality for instantiating another browser window. This function can be invoked as a logical change to a property of the object. The second browser window can be referenced through the original browser window (i.e., object) or designated as a new live object.

[0044] An information provider 108 or other entity preferably updates a live object at a client 114 via an update message. In general, an update message identifies the live object and, if necessary, the property of the live object, and contains data for updating the property. In one embodiment, the data may be the actual value for the property or executable code for causing the object's property to be updated. For example, the data may be a simple numerical or textual value, e.g., "4," to which the property should be set,

and/or the data may be JAVASCRIPT[®] code or a call to a JAVASCRIPT[®] function at the client that effects the desired change to the property of the object.

[0045] The update message preferably implicitly or explicitly identifies a handler at the client 114 for use in updating the live object's property. In one embodiment, the client 114 utilizes a default handler when the message implicitly specifies the handler (e.g. when the message does not identify a specific handler). In one embodiment, if the update message specifies the actual value for the property, a default handler generates JAVASCRIPT[®] code for changing the property to the specified value. If the data in the update message are JAVASCRIPT[®] code, the default handler does not perform any processing of the code. In either case, the default handlers preferably use LiveConnect to execute the JAVASCRIPT[®] code in a Java Virtual Machine (JVM) 122 at the client 114 and thereby update the property of the live object.

[0046] For certain objects and/or data types, the default handlers are not appropriate. In these cases, the message preferably explicitly identifies a handler for performing the update. For example, the message may explicitly specify a function to call on the data or the message may explicitly identify the environment in which the data should be executed. For example, the data in the update message may include code for execution by a software "plug-in" such as MACROMEDIA FLASH[®] and the message may explicitly identify FLASH as the handler.

[0047] The information provider 108 preferably designates an object as "live" by including a unique identifier for the object, the object ID, in the web page or other data representation provided to the client 114. In one embodiment, the information provider

108 encodes the object ID in an object's corresponding HTML "ID" attribute using the following HTML expression:

ID = "elementIdentifier,"

where "elementIdentifier" is the object ID and is preferably a string. The string can encode any information desired by the information provider 108 or other entity establishing the object ID and in one embodiment is a simple textual and/or numeric identifier. In one embodiment, the information provider 108 begins the object ID with a predefined token, such as "Bang\$," in order to distinguish live objects from other objects that happen to have defined ID attributes. For example, an object can have the object ID "Bang\$elementIdentifier."

[0048] In the preferred embodiment, each information provider 108 optionally encodes a unique information provider ID in its object IDs in order to prevent naming collisions between the object IDs of different information providers. In one embodiment, the information provider ID is a textual and/or numeric identifier. The information provider 108 may specify the information provider ID and the object ID as part of a hierarchical namespace. For example, in one embodiment objects are named as follows: "\$namespace1\$[namespace2\$...\$namespaceN\$]objectId," where "\$namespace1" is the information provider ID and the "\$" operates as the name separator and defines additional optional levels of a namespace hierarchy. One embodiment of the system 100 supports typical directory services functionality. For example, two dollar sign characters appearing together, "\$\$", refers to the top level of the namespace hierarchy.

[0049] Thus, the object ID for a live object is preferably formed from a combination of the predefined token, the information provider ID namespace, and a value assigned by the information provider 108. For example, the object ID for a live object representing the real time price of a stock having the symbol "BANG" might be:

"Bang\$\$informationProviderID\$equities\$realtime\$bang." In this example, "Bang\$" is the predefined token that signifies a live object, "\$informationProviderID" is the ID identifying the information provider, "\$equities\$realtime\$" defines levels of a namespace hierarchy, and "bang" identifies the specific object.

[0050] In some embodiments and situations, the object ID utilizes relative names. For example, an information provider 108 referring to its own object IDs is implicitly in its own namespace. Accordingly, the information provider 108 does not need to include the information Provider ID in the object IDs it utilizes internally. In one embodiment, the information provider ID is not explicitly encoded into the object ID. Instead, the information provider ID is encoded elsewhere in the web page in order to provide scope to the page's object IDs.

[0051] In one embodiment, the object ID identifies a point (i.e., a node in a tree) in a Document Object Model (DOM) representation of a web page or other document at the client 114. The DOM is a platform- and language-neutral interface that represents a document as a hierarchy of objects. The DOM also provides an interface that allows programs and scripts to dynamically access and update properties of the objects. Object properties can be inherited by descendent objects.

[0052] In this embodiment, the client 114 preferably executes an update message in the context of the specified point in the DOM representation. The update may specify a change to a property of the object at the identified point. The update also may specify a change to a parent or descendent of the object at the identified point. In each case, the update is executed relative to the specified point in the DOM representation. In one embodiment, points in the DOM representation specify how to update properties of live objects located at those points. Thus, the same update may be interpreted differently depending upon the identified live object's location in the DOM representation.

[0053] For example, assume there is an object in the DOM representation identified as "window.document.frame[3].ObjectID." Also assume that the object has an "innerText" property located at "window.document.frame[3].ObjectID.innerText" that specifies the text displayed by the object. An update message can change the text displayed by the object by specifying "ObjectID" and the new value for the innerText property.

[0054] An advantage of utilizing object IDs to specify objects is that the information provider 108 or other entity providing the update message can access and change properties of objects without knowing the object's actual location in the DOM representation. Indeed, the object may be in different locations in different DOM representations and/or in multiple locations in the same DOM representation. In any of these cases, the update message will change the specified properties of all of the objects having the given object ID.

[0055] Depending upon the particular embodiment of the environment 100, the information provider 108 and/or the dynamic content provider 116 provides update messages to the routing network 110. The dynamic content provider 116 is preferably a conventional computer system operated by an entity that provides real-time information, such as stock prices and/or sports scores. In one embodiment, the information provider 108 receives updated properties for the live objects from the dynamic content provider 116 or another source (or generates the updated properties internally). Then, the information provider 108 sends an update message specifying the object ID and the change to the object property to the routing network 110. In this embodiment, the dynamic content provider 116 may be absent from the environment 100.

[0056] In another embodiment, the dynamic content provider 116 provides the object IDs for live objects to one or more information providers 108 and the information providers 108 distribute the live objects to the clients 114. Then, the dynamic content provider 116 sends messages specifying the changes to the properties of the live objects to the routing network 110. For example, the dynamic content provider 116 distributes an object ID associated with the score of a particular baseball game to the information providers 108. Then, the dynamic content provider 116 sends a message specifying the object ID and an update to a property of the object that controls the displayed score of the particular baseball game to the routing network 110. These two embodiments are not mutually exclusive and, therefore, some updates may be provided to the routing network 110 by the information provider 108 while others are provided by the dynamic content provider 116.

[0057] The client 114 is a device that retrieves web pages 118 and/or other information from the server 112. In one embodiment, the client 114 is a conventional personal computer used by a person to access information on the Internet. In alternative embodiments, the client 114 is a different consumer electronic device having Internet connectivity, such as an Internet-enabled television, a cellular telephone, a personal digital assistant (PDA), a web browsing appliance, etc. The client 114 preferably, but not necessarily, has an associated display device.

[0058] The client 114 preferably executes a web browser 120, such as MICROSOFT INTERNET EXPLORER®, for retrieving web pages and displaying them on the display device. In embodiments where the client receives data representations from the server 112 other than conventional web pages, the web browser 120 does not necessarily share similarities with conventional web browsers. Preferably, the web browser 120 contains a JVM 122 for executing JAVA® applets and/or scripts. The web browser 120 also preferably contains Dynamic HTML capabilities, such as support for JAVASCRIPT® (or another scripting language, such as VBScript) and the Document Object Model (DOM), and enables communications between JAVA® and the scripting languages. In one embodiment, the web browser 120 supports the LiveConnect standard for enabling communication between JAVA® applets and scripts written in the supported scripting languages. The web browser 120 can also be extended through software plug-ins such as MACROMEDIA FLASH®, REAL NETWORKS REALPLAYER®, and/or APPLE QUICKTIME®. In alternative embodiments, the functionality of the JVM 122 and/or other aspects of the web browser 120 are provided by one or more other functional units

within the client 114. The term "module" is used herein to refer to software computer program code and/or any hardware or circuitry utilized to provide the functionality attributed to the module. The web browser 120 and JVM 122 are examples of modules in the client 114.

[0059] In some embodiments, the client 114 does not necessarily have a display device, web browser 120 and/or other components associated with a typical consumer device. The client 114, for example, may be a dedicated purpose device having certain aspects of web connectivity such as an embedded HTTP client in a web-enabled appliance or in a controller for an automobile, audio-visual equipment, or some other device.

[0060] A web page 118 provided from the server 112 to the client 114 preferably includes instructions for enabling the live objects on the web page. The instructions cause the client 114 to automatically and transparently (i.e., without user interaction) contact the routing network 110 and download an activation module 124 for activating the live objects. In one embodiment, the instructions comprise a URL specifying the location of the activation module 124 at the routing network 110. In an alternative embodiment, the client 114 obtains the activation module 124 from the server 112 or another source.

[0061] The activation module 124 preferably contains JAVA[®] instructions for execution by the JVM 122. However, alternative embodiments of the module 124 may encode the instructions in the web page 118 and/or the activation module 124 using different languages and/or techniques. For example, the instructions and/or activation

module 124 can be embedded in the web browser 120 or operating system, either as native code or as plug-ins. In these alternative embodiments, the web browser 120 does not have to download the activation module 124 from an external source.

[0062] The activation module 124 preferably registers object IDs from the web page 118 downloaded by the client 114 with the routing network 110 and updates the live objects in response to update messages received from the network. The routing network 110 records the registrations in the registry 125. The client's registrations preferably remain in effect as long as the client is displaying the associated web page 118, although other embodiments of the system 100 may use different criteria for determining when to terminate the client's registrations.

[0063] FIG. 2 is an interaction diagram illustrating interactions among the server 112, information provider 108/dynamic content provider 116 (generically referred to as an "input source 210"), client 114, and the routing network 110 to update a property of a live object. Initially, the client 114 sends 212 a web page request to the server 112. In response, the server 112 provides 214 to the client 114 the web page containing or otherwise identifying the one or more live objects. Instructions encoded in the web page preferably cause the client 114 to transparently request 216 the activation module 124 from the routing network 110. In response, the routing network 110 sends 218 the activation module 124. The client 114 executes 220 the activation module 124, which identifies the object IDs of the live objects at the client and registers 222 the object IDs with the routing network 110. The routing network 110 updates 223 its registry to identify the object IDs for which the client 114 has registered.

[0064] At some point, the input source 210 sends 224 an update message to the routing network 110 in order to change a property of a live object at the client 114. In one embodiment, the message from the input source 210 to the routing network 110 contains only a single object ID and an update to a property of the identified object. In another embodiment, the message contains multiple object IDs and the corresponding property updates. In this latter embodiment, the message may have an associated "Batch ID" that identifies the message as having multiple object IDs and updates. Preferably, the information provider 108 can include a batch ID in a web page 118 in the same manner as including an object ID. Likewise, the client 114 can preferably register for a batch ID with the routing network 110 in the same manner as an object ID. In fact, the batch ID can be the same as the object ID so that the client 114 registers for both batch and non-batch messages by registering one ID. Alternatively, separate procedures can be established for registering batch messages. The client 114 preferably processes the component messages of a batch as if each message were delivered separately.

[0065] The routing network 110, in turn, routes 226 the message to each client 114 that has registered for the specified object ID, preferably by utilizing standard Internet communications protocols, such as IP addresses, etc. The activation module 124 at the client 114 processes the message and updates 228 the property of the identified live object. If live objects having the same object ID appear in multiple locations at the client 114 (e.g., at multiple locations on a web page being displayed at the client), the activation module 124 preferably updates each of the live objects having the specified ID. As a result, the routing network 110 allows live objects at the client 114 to be dynamically

updated. Preferably, this routing and updating happens quickly enough to be considered “real-time” for the purposes of the input source 210.

[0066] This update process, indicated within the dashed box 230 in FIG. 2, can repeat an indefinite number of times and is fully asynchronous as to the information provider 210 and client 114. For example, the input source 210 may send regular update messages to the routing network 110 as the score of a sporting event changes or a stock price fluctuates, but may stop sending update messages once the sporting event ends or stock market closes. When the client 114 ends the display of a web page containing the live object, or otherwise no longer desires to receive update messages, the client preferably closes 232 the connection with the routing network 110. The routing network 110, in turn, updates 234 the registry 125 to remove the client’s object registrations. In another embodiment, the client 114 sends messages to the routing network 110 that selectively register and/or de-register the client from one or more objects yet leaves the connection open in order to receive update messages pertaining to other objects.

[0067] FIG. 3 is a high-level diagram graphically indicating the many-to-many mapping performed by the routing network 110. Multiple input sources (labeled 210A-C) send update messages to the routing network 110. Each update message preferably specifies at least one object ID and an update to a property of the identified object. The routing network 110, in turn, selectively routes the update messages to the clients 114 that have registered for the given object ID from the given input source 210. In FIG. 3, assume for example that clients 312A and 312B have registered for a given object ID while the other clients have not registered for the object ID. Accordingly, the routing

network 110 routes the update message to clients 312A and 312B, but does not route the message to clients 312C-312H.

[0068] FIG. 4 illustrates an example of the capabilities of the dynamic content routing network 110. FIG. 4 illustrates two different web pages 410, 412 containing sports scores. Although the web pages are formatted differently, each page contains the same scores for two professional football games and two professional baseball games. Web page 410 contains all four games under the heading "Local Sports Scores" while web page 412 contains the baseball games under the heading "Baseball Scores" and the football games under the heading "Football Scores."

[0069] There are various ways to internally represent the games and scores in the web pages using live objects. In one embodiment, a "game" object is defined having properties for the two teams involved in the game and the score associated with each team. The game object is placed at a selected position in the web page and the properties of the object cause the information about the game to be displayed on the page. In another embodiment, "team" and "score" objects are defined, with the team object having a property defining the name of a team and the score object having a property defining a score. In this second embodiment, the team and score objects are placed at selected locations on the page so that the proper teams and scores are aligned when the page is rendered. In yet another embodiment, an object is defined having properties for the name of one team and a score associated with that team. Then, pairs of the objects are placed in the page in the proper alignment to indicate the games and scores. In another embodiment, an object is defined having properties specifying names of two teams and a

separate object is defined having properties specifying two scores. In this last embodiment, the two objects are placed in the page so that the names of the teams align with the associated scores. Obviously, additional variations of these representations are possible.

[0070] Assume for the example of FIG. 4 that the names of teams in a game are specified by a "names" object having properties for the two team names and the scores in the game are specified by a "scores" object having properties for two scores. In web page 410, a names object 414 having properties set to identify the "SF 49ers" and the "STL Rams" is located directly under the "Local Sports Scores" heading. A scores object 416 having a property set to identify the score of the game as "42" to "7" is directly to the right of the names object 414. In web page 412, the properties of the second names object 418 identify the same game using slightly different terminology: "SF" and "STL." However, this names object 418 is aligned with the same scores object 416 as is utilized in web page 410.

[0071] Thus, the same scores object 416 is utilized in different positions in each web page 410, 412. In order to update the score of the San Francisco 49ers vs. St. Louis Rams football game on both web pages, the input source 210 simply sends an update message to the routing network 110 specifying the object ID for the scores object 416 and the update to the score property. The routing network 110 routes the update message to the appropriate clients 114, and the clients update the appropriate score regardless of the particular page layout.

[0072] The input source 210, i.e., the information provider 108 and/or dynamic content provider 116 can use a variety of tools to generate the update messages. FIG. 5 is a block diagram illustrating an input source 210 and the tools available to it for generating the update messages. Other tools can be utilized in addition to or instead of the ones described herein.

[0073] Preferably, the tools allow the input source 210 to access an application programming interface (API) provided by the routing network 110 for accepting messages. In one embodiment, the messages sent by the input source 210 are in the same format as utilized by the activation module 124 at the client 114. In an alternative embodiment, the messages provided to the routing network 110 are in a different format and the routing network translates the messages into the format utilized by the activation module 124.

[0074] In one embodiment, the input source 210 utilizes a data pump module 510 to access the API. The data pump module 510 reads an extensible markup language (XML) file containing one or more object IDs and the new values for the identified objects at regular intervals and automatically generates API calls that send messages representing changes to object properties to the routing network 110. In another embodiment, the data pump module 510 is event-driven and reads the XML file in response to a change in the file or some other occurrence.

[0075] In another embodiment, the input source 210 utilizes a director console module 512 to access the API. Preferably, the director console module 512 presents an administrator with a graphical interface displaying the contents of the web page 118. For

example, the administrator may use the director console 512 to edit textual data, images, and/or any objects or properties of objects on the web page. After editing, the administrator uses a “send update” button or similar technique to cause the director console module 512 to send messages for the changed objects and properties to the routing network 110 via the API.

[0076] In another embodiment, the information provider 108 and dynamic content provider 116 work together as the input source 210 by using a content management system module 514 to access the API. Preferably, the content management system module 514 resides at the information provider 108 and receives object property updates from the dynamic content provider 116. The content management system module 514 preferably updates the properties of the live objects in the web page 118 stored at the server 112 and also sends messages for the changed properties to the routing network 110. In this manner, the web page 118 at the server 112 and the web page displayed at the client 114 are updated almost simultaneously. In one embodiment, the dynamic content provider 116 sends the update messages to the routing network 110 instead of to the information provider 108. Embodiments of the system 100 can also utilize any combination of the content management techniques described herein.

[0077] For example, the tools described above can generate a message having the following code for updating the text displayed by a score object to “2”:

```
LiveObject score = new LiveObject("Bang$homeScoreID");  
score.setProperty("innerText", "2").
```

This code sets the innerText property of the object having object ID

“Bang\$homeScoreID” to “2.” The tools use the API to pass this message to the routing network 110.

[0078] Turning now to the actions performed at the client 114, FIG. 6 is a flow chart illustrating the steps performed by an embodiment of the activation module 124. Those of skill in the art will recognize that different embodiments may perform the steps of FIG. 6 in different orders. The activation module 124 generally performs three functions: register object IDs with the routing network 110, handle messages received by the client 114 from the network in order to update the properties of live objects, and control communications between the client and the network.

[0079] In order to register object IDs, the activation module 124 preferably parses 610 the web page 118 received from the server 112 and identifies the object IDs of the live objects. In an alternative embodiment, the activation module 124 identifies only a subset of the object IDs, such as the IDs of only live objects that are currently being displayed by the web browser 120. Alternatively, a list of object IDs may be pre-encoded in the web page in addition to the objects themselves, thereby enabling easy identification by the activation module 124. In yet another embodiment, a user of the client 114 selects the object IDs to register.

[0080] The activation module 124 preferably opens 612 a connection between the client 114 and the routing network 110. The activation module 124 can open 612 this connection before or after the activation module receives and/or parses the web page 118. In some cases, the client 114 is located behind a firewall that puts a restriction on the

types of connection requests the client can make. A firewall might, for example, block all non-HTTP traffic. For this reason, the activation module 124 preferably wraps the connection request in an HTTP header in order to get the request to the routing network 110 through the firewall.

[0081] The activation module 124 uses the connection between the client 114 and routing network 110 to register 614 the object IDs by communicating to the routing network 116 a vector (e.g., a list or array) containing the identified object IDs. In order to accomplish this task through the firewall, the activation module 124 preferably puts the vector into a string, referred to as "object data," and then preferably creates an HTTP message to communicate the object data to the routing network 110. A schematic example is as follows:

```
POST / HTTP/1.1\r\n
Content-Length: <length of object data> \r\n
\r\n
<object data>
```

where <object data> is the object ID list. When the routing network 110 receives such an HTTP request, it extracts the object data and updates the registry 125 to indicate that the client 114 has registered for the identified objects.

[0082] If the web browser 120 loads 616 a new page, or otherwise terminates display of the objects on the initial page, the activation module 124 associated with the initial web page preferably terminates 618 the client's connection with the routing network 110. Those of skill in the art will recognize that this termination 618 can occur asynchronously with the other steps illustrated in FIG. 6. Thus, the location of steps 616 and 618

represents only one possible place in the sequence of steps where the termination may occur.

[0083] If the connection is not terminated, the activation module 124 preferably waits until it receives 618 a message from the routing network 110 specifying an object ID and an update to a property of the identified object. In one embodiment, this message is received as HTTP data. Upon receipt of the message, the activation module 124 preferably extracts 620 the object ID and update from the HTTP data. Then, the activation module 124 updates 622 the property of the identified object, or causes the object to be updated, as specified by the message.

[0084] The sequence of receiving messages 618, extracting data 620, and updating objects 622 is preferably repeated until a new page is loaded 616 or the connection with the routing network 110 is otherwise terminated. Although not shown in FIG. 6, in certain circumstances, such as when a user action with respect to the web page 118 activates a new live object, the activation module 124 may register new object IDs with the routing network 110 without first downloading and parsing a new page. In one embodiment, if the newly-loaded page contains live objects, then the process of downloading the activation module 124 and updating the objects as described by FIG. 6 is repeated. In an alternative embodiment, the activation module 124 remains active at the client 114 and, therefore, the client does not re-download the activation module from the routing network 110. Instead, the already-present activation module 124 performs the live-enabling process on the new page.

[0085] FIG. 7 is a block diagram illustrating a lower-level view of the routing network 110 according to an embodiment of the present invention. Those of skill in the art will recognize that there are many alternative ways to implement the functionality of the routing network 110. FIG. 7 illustrates multiple input sources (labeled 710A-D) representative of sources providing messages to the routing network 110, such as an information provider 710A and a dynamic content provider 710B. FIG. 7 also illustrates multiple clients (labeled 712A-F) representative of the many clients in communication with the routing network 110 at any given instant.

[0086] Internally, the routing network 110 is preferably divided into one or more clusters 714. In FIG. 7, the routing network 110 has three clusters 714A, 714B, 714C, although the number of clusters can vary depending upon the processing needs of the network. An input-side global load balancer 716 preferably routes messages from the input sources 710 to the clusters 714. Similarly, a client-side global load balancer 718 preferably routes connection requests from the clients 712 to the clusters 714. The load balancers 716, 718 are designed to ensure that load is distributed among the clusters 714 according to a predetermined heuristic. For example, the load may be distributed evenly among the clusters 714 or a more powerful cluster may be distributed a majority of the load. In one embodiment, one load balancer performs the functions of the input-side 716 and client-side 718 load balancers and utilizes conventional Domain Name System- (DNS-) based load balancing.

[0087] Each cluster 714, of which cluster 714A is representative, preferably contains an input-side cluster load balancer 720A and a client-side cluster load balancer 722A.

The cluster load balancers 720A, 722A function similarly to the corresponding global load balancers 716, 718 in that the input-side cluster load balancer 720A balances and routes incoming messages among one or more gateways 724A and the client-side cluster load balancer 722A balances and routes incoming connection requests among one or more nodes 726A and application servers 728A.

[0088] In one embodiment, the functionality of the two client-side cluster load balancers 720A, 722A is provided by one component. This single-component load balancer initially determines whether an incoming request is from an input source 710 seeking to send a message to a gateway 724A, a client 712 seeking a connection to a node 726A, or a client seeking a connection to an application server 728A. Then, the load balancer routes the messages/connection requests among the gateways 724A, nodes 726A, and application servers 728A within the cluster 714. In one embodiment, the single-component load balancer provides layer seven load balancing (i.e., load balancing at the application layer). Preferably, the load balancing for the nodes 726A and application servers 728A are performed by the same component since, for security reasons, most client web browsers only permit an application (e.g., the activation module 124) to transparently connect to the location from which the application was downloaded.

[0089] Alternative embodiments of the routing network 110 may combine the global 716, 718 and cluster 720A, 722A load balancers and/or incorporate the functionality of the load balancers into different components within or outside of the clusters 714. In addition, alternative embodiments may omit one or more of these load balancers. For

example, having different clusters 714 serve different customers might obviate the need for the global load balancers 716, 718.

[0090] The gateways 724A in the cluster 714 receive the messages from the input sources 710 and direct the messages to the appropriate node or nodes 726A. In one embodiment, each gateway 724A maintains a persistent TCP connection to every node 726 in every cluster 714 and directs every message to every node. Therefore, although a gateway 724A is located inside a cluster 714A and receives connections via the cluster's input-side load balancer 720A, the gateway's scope spans the entire routing network 110. This broad scope allows messages from any input source to reach any client 712.

[0091] In an alternative embodiment of the routing network 110, each gateway 724 maintains a persistent TCP connection to all nodes 426 in the same cluster 714 and at least one connection to at least one gateway in each of the other clusters. This embodiment reduces the number of simultaneous TCP connections maintained by each gateway 724. In another alternative embodiment, each cluster 714 also includes a gatekeeper (not shown in FIG. 7) that maintains connections with the gateways 724 of other clusters. A gateway 724 forwards messages to the gatekeeper, which then distributes the messages to the gateways of other clusters 714.

[0092] Since a gateway 724 does not control the rate at which it receives messages from input sources 710, it is possible for the gateway to receive messages faster than it can process them (i.e., send the messages to the nodes). Therefore, each gateway 724 preferably maintains a queue 730 of messages that have been received but not yet processed in order to avoid losing messages. In one embodiment, the gateway 724 drops

messages if the queue 730 becomes too long. In another embodiment, the gateway 724 utilizes priorities assigned to certain messages or input sources to determine which messages to drop.

[0093] The nodes 726 preferably transmit messages received from the gateways 724 to the clients 712 that have registered in the object IDs identified by the messages. If no clients 712 have registered the object ID specified by a message, the node preferably ignores the message. A node 726 preferably maintains an instance of the registry 125 as a hash table 732 containing the object IDs registered by clients 712 connected to the node. In one embodiment, the hash table 732 associates each object ID with a linked list containing one entry for each client 712 that has registered for that object ID. Each entry in the linked list preferably contains a pointer to a socket representing the connection to the corresponding client 712. As is known in the art, the pointer to the socket, typically called a "file descriptor," represents an address to which the node can write in order to send the message to the corresponding client. Preferably, the node 726 adds to the hash table 732 and/or linked list every time a client 712 registers an interest in an object and deletes the corresponding entry from the hash table and/or linked list when the client disconnects from the node or otherwise indicates that it is no longer interested in a particular object.

[0094] Alternative embodiments of the present invention utilize other data structures in addition to, or instead of, the hash table 732 and linked list, and/or may utilize different data within the data structures. For example, one embodiment of the routing network 110 has a hierarchy of nodes within each cluster 714. Different nodes in the hierarchy may

handle messages received from certain input sources 210, or process messages sent to different clients 712. In this embodiment, the linked lists may point to nodes lower in the hierarchy, instead of to sockets leading to the clients 712. Another embodiment lacks the node hierarchy, yet assigns certain nodes to certain input sources 210 or clients 712.

[0095] The application server 728 within each node 714 preferably serves the activation module 124 to the clients 712 in response to client requests. In addition, the application server 728 serves any other modules that may be required or desired to support the environment 100. In an alternative embodiment of the routing network, a single application server 728 fulfills all of the client requests. This application server 728 may be within a certain cluster 714 or independent of the clusters. However, this single-application-server embodiment is less desirable because it lacks redundancy.

[0096] Preferably, the routing network 110 utilizes conventional single-processor computer systems executing the Linux operating system (OS). Preferably, each component of the routing network 110 is implemented by a separate, dedicated computer system in order to enable the separate optimization of the components. The input/output (I/O) functionality of the OS is preferably enhanced through the use of a non-blocking OS package such as NBIO available from the University of California, Berkeley, California. Based on the assumption that connections with the nodes 728 are long-lived, the OS is preferably configured to not allocate resources toward monitoring idle connections. Instead, the well-known /dev/poll patch is preferably applied to the OS in order to provide advanced socket polling capabilities.

[0097] Moreover, the TCP/IP stack in the OS is preferably optimized in order to quickly output messages. In one embodiment, the retransmit timer in the stack is reduced from 200 ms to 50 ms. This timer determines how long the stack waits for an acknowledgement (ack) that a sent packet was received. Due to the way the Linux kernel implements the retransmit timer, the kernel will not send pending outbound packets (even if the ack has been received) until the initial retransmit timer has expired. Reducing the retransmit value minimizes the effect of this delay. If an ack is not received before the retransmit timer expires, an embodiment of the present invention increases the retransmit value for the affected TCP connection and the unacknowledged packet is retransmitted. In addition, the TCP/IP stack preferably utilizes Nagle's algorithm functionality to concatenate a number of small messages into a larger message, thereby reducing the number of packets sent by the routing network 110.

[0098] FIG. 8 is a flow chart illustrating steps performed by a node 726 in a cluster 714 to perform object-based routing of a message received from an input source via the gateway 724. Initially, the node 726 receives 810 the message from an input source 710. The node 726 extracts 812 the object ID from the message. In addition, the node 726 optionally transforms 814 the message to convert it from the format utilized by the input source 710 into the format utilized by the activation module 124 at the client 712. As described above, in one embodiment the message format utilized by the input source 710 is identical to the message format utilized by activation module 124. In this embodiment, therefore, the node 726 does not need to transform the message. In an alternative embodiment wherein the input source 710 and activation module 124 utilize different

message formats, the node 726 preferably transforms the message. The node 726 looks up 816 the hash table entry corresponding to the extracted object and information provider IDs to determine the linked list of clients 712 that have registered in the object referenced by the message. Finally, the node 726 transmits 818 the message to each of the registered clients 712. In an alternative embodiment, the node 726 optionally transforms the message after, as opposed to before, looking up the registered clients in the hash table. Transforming the message at this latter stage enables the node 726 to transform the message according to the specific requirements of the registered clients 712.

[0099] The above description is included to illustrate the operation of the preferred embodiments and is not meant to limit the scope of the invention. The scope of the invention is to be limited only by the following claims. From the above discussion, many variations will be apparent to one skilled in the relevant art that would yet be encompassed by the spirit and scope of the invention.